

TUGAS AKHIR

**PREDIKSI POPULARITAS GIM INDI PADA
PLATFORM STEAM MENGGUNAKAN ALGORITMA
CATBOOST DAN SHAPLEY ADDITIVE
EXPLANATIONS (SHAP)**



MOHAMMAD TEDDY SYAMKALLA

20110013

**PROGRAM STUDI SAINS DATA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2024**

TUGAS AKHIR

PREDIKSI POPULARITAS GIM INDI PADA PLATFORM STEAM MENGGUNAKAN ALGORITMA CATBOOST DAN SHAPLEY ADDITIVE EXPLANATIONS (SHAP)

PREDICTING THE POPULARITY OF INDIE GAMES ON STEAM PLATFORM USING CATBOOST ALGORITHM AND SHAPLEY ADDITIVE EXPLANATIONS (SHAP)

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Sains Data



MOHAMMAD TEDDY SYAMKALLA

20110013

**PROGRAM STUDI SAINS DATA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2024**

LEMBAR PERSETUJUAN PEMBIMBING

**PREDIKSI POPULARITAS GIM INDI PADA
PLATFORM STEAM MENGGUNAKAN ALGORITMA
CATBOOST DAN SHAPLEY ADDITIVE
EXPLANATIONS (SHAP)**

**PREDICTING THE POPULARITY OF INDIE GAMES
ON STEAM PLATFORM USING CATBOOST
ALGORITHM AND SHAPLEY ADDITIVE
EXPLANATIONS (SHAP)**

Dipersiapkan dan Disusun oleh
MOHAMMAD TEDDY SYAMKALLA
20110013

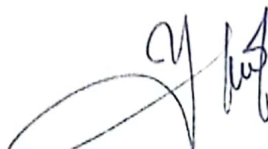
**Fakultas Informatika
Institut Teknologi Telkom Purwokerto
Pada Tanggal :11 Januari 2024**

Pembimbing Utama,



Siti Khomsah S.Kom., M.Cs.
NIDN : 0517108101

Pembimbing Pendamping,



Yohani Setiya Rafika Nur, S.Kom., M.Kom.
NIDN : 0627099501

**PREDIKSI POPULARITAS GIM INDI PADA
PLATFORM STEAM MENGGUNAKAN ALGORITMA
CATBOOST DAN SHAPLEY ADDITIVE
EXPLANATIONS (SHAP)**

**PREDICTING THE POPULARITY OF INDIE GAMES
ON STEAM PLATFORM USING CATBOOST
ALGORITHM AND SHAPLEY ADDITIVE
EXPLANATIONS (SHAP)**

Disusun oleh
MOHAMMAD TEDDY SYAMKALLA
20110013

Telah Diujikan Dan Dipertahankan Dalam Sidang Ujian Sidang Akhir
Pada Kamis, Tanggal 18, Januari 2024

Penguji I,



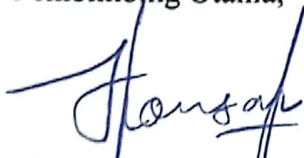
Auliya Burhanuddin S.Si., M.Kom.
NIDN : 0630058202

Penguji II,




Umami Athiyah, S.Kom., M.Kom.
NIDN : 0621129001

Pembimbing Utama,



Siti Khomsah S.Kom., M.Cs.
NIDN : 0517108101

Pembimbing Pendamping,



Yohani Setiya Rafika Nur, S.Kom., M.Kom.
NIDN : 0627099501

Dekan,



Auliya Burhanuddin S.Si., M.Kom
NIK : 19820008

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Mohammad Teddy Syamkalla**
NIM : **20110013**
Program Studi : **S1 Sains Data**

Menyatakan bahwa Tugas Akhir dengan judul berikut:

**PREDIKSI POPULARITAS GIM INDI PADA PLATFORM STEAM
MENGUNAKAN ALGORITMA CATBOOST DAN SHAPLEY
ADDITIVE EXPLANATIONS (SHAP)**

Dosen Pembimbing Utama : Siti Khomsah S.Kom., M.Cs

Dosen Pembimbing Pendamping : Yohani Rafika Nur, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Institut Teknologi Telkom Purwokerto maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan, dan penelitian Saya Sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam Karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggungjawab Saya, bukan tanggungjawab Institut Teknologi Telkom Purwokerto.
5. Pernyataan ini Saya buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima Sanksi Akademik dengan pencabutan gelar yang sudah diperoleh serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Purwokerto, 11 Januari 2024,

Yang Menyatakan,



The image shows a handwritten signature in black ink over a 2000 Rupiah postage stamp. The stamp features the Garuda Pancasila emblem and the text 'REPUBLIK INDONESIA', '2000', 'MELUKAI', and 'TEMPER'. The serial number 'EA1FCAKX757288237' is visible at the bottom of the stamp.

Mohammad Teddy Syamkalla

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT yang Maha kuasa, atas seluruh curahan rahmat dan hidayahNya sehingga peneliti mampu menyelesaikan tugas akhir dengan judul “PREDIKSI POPULARITAS GIM INDI PADA PLATFORM STEAM MENGGUNAKAN ALGORITMA CATBOOST DAN SHAPLEY ADDITIVE EXPLANATIONS (SHAP)” tepat waktu. Tugas akhir ini disusun Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Sains Data Fakultas Informatika Institut Teknologi Telkom Purwokerto. Dalam penyelesaian penelitian ini, peneliti banyak memperoleh bantuan dari berbagai pihak baik itu secara langsung maupun secara tidak langsung. Untuk itu peneliti ingin mengucapkan terima kasih atas bimbingan dan dukungan yang luar biasa diberikan kepada :

1. Allah SWT yang senantiasa memberikan petunjuk dan kekuatan kepada peneliti dalam menyelesaikan tugas akhir ini.
2. Kedua orang tua peneliti, Teguh Suryahadi dan Suratin Puspita Dini, yang telah memberi kasih sayang, doa, nasehat, serta usaha yang tidak akan mampu peneliti balas hingga kapanpun. Merupakan anugerah terbesar dalam hidup peneliti. Semoga penulis mampu menjadi anak yang dapat kalian banggakan.
3. Ibu Dr. Tenia Wahyuningrum, S.Kom., M.T. selaku Rektor Institut Teknologi Telkom Purwokerto.
4. Bapak Auliya Burhanuddin S.Si., M.Kom. selaku Dekan Fakultas Informatika Institut Teknologi Telkom Purwokerto, sekaligus dosen penguji 1 yang telah meluangkan waktu ditengah kesibukan beliau untuk menguji kelayakan Tugas Akhir yang telah diselesaikan oleh peneliti.
5. Ibu Siti Khomsah S.Kom., M.Cs. selaku Kaprodi Sains Data, Dosen Wali sekaligus Dosen Pembimbing Utama yang telah meluangkan waktu ditengah kesibukan beliau, memberikan kritik, masukan dan arahan kepada peneliti dalam menempuh studi di Prodi Sains Data hingga proses penyelesaian Tugas Akhir ini.

6. Ibu Ummi Athiyah, S.Kom., M.Kom. selaku dosen penguji 2 yang telah meluangkan waktu ditengah kesibukan beliau untuk menguji kelayakan Tugas Akhir yang telah diselesaikan oleh peneliti.
7. Ibu Yohani Setiya Rafika Nur, S.Kom., M.Kom. selaku dosen pembimbing pendamping yang telah meluangkan waktu ditengah kesibukan beliau, memberikan kritik, masukan dan arahan kepada peneliti dalam proses penyelesaian Tugas Akhir ini.
8. Seluruh jajaran staff dosen di Fakultas Informatika yang telah memberi ilmu pengetahuan yang sangat berharga selama peneliti menempuh pendidikan di Fakultas Informatika Institut Teknologi Telkom Purwokerto.
9. Para adik peneliti, Yusuf Sanjaya, Angger Wijaya, dan Anggun Sekar Kinami. Semoga penulis dapat menjadi contoh seorang kakak yang baik untuk kalian.
10. Seluruh teman seperjuangan yang tidak mampu peneliti sebutkan seluruhnya pada bagian kata pengantar ini. Untuk semua pengalaman bermanfaat yang peneliti rasakan selama menempuh pendidikan di Fakultas Informatika Institut Teknologi Telkom Purwokerto.

Semua bantuan dan dukungan ini menjadi pilar lancarnya proses penelitian. Sekali lagi, terima kasih kepada semua pihak yang telah berperan dalam kesuksesan penyelesaian tugas akhir ini. Semoga hasil penelitian ini bermanfaat bagi seluruh pihak khususnya pada pengembangan ilmu pengetahuan di masa yang akan datang.

Purwokerto, 11 Januari 2024



Mohammad Teddy Syamkalla

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	iii
LEMBAR PENGESAHAN TUGAS AKHIR	iv
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
ABSTRAK.....	xiv
ABSTRACT.....	xv
BAB 1	1
PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	4
1.3 Pertanyaan Penelitian.....	4
1.4 Tujuan	4
1.5 Batasan Masalah	4
1.6 Manfaat Penelitian	4
BAB 2	6
TINJAUAN PUSTAKA DAN LANDASAN TEORI	6
2.1 Tinjauan Pustaka.....	6
2.2 Landasan Teori.....	17
2.2.1 Gim Indi.....	17
2.2.2 <i>Machine Learning</i>	18
2.2.3 <i>Exploratory Data Analysis (EDA)</i>	19
2.2.4 <i>CatBoost (Categorical Boosting)</i>	19
2.2.5 Matriks Konfusi	22
2.2.6 Kurva <i>AUC (Area Under the Curve)</i>	23
2.2.7 <i>Shapley Additive Explanations (SHAP)</i>	24
BAB 3	26
METODOLOGI PENELITIAN.....	26
3.1 Objek dan Subjek Penelitian.....	26

3.2	Alat dan Bahan Penelitian.....	26
3.2.1	Alat Penelitian.....	26
3.2.2	Bahan Penelitian	26
3.3	Diagram Alir Penelitian	27
3.3.1	Identifikasi Masalah.....	27
3.3.2	Studi Literatur	27
3.3.3	Pengumpulan Data	28
3.3.4	<i>Exploratory Data Analysis (EDA)</i>	29
3.3.5	<i>Preprocessing Data</i>	30
3.3.6	<i>Feature Engineering</i>	32
3.3.7	<i>Modelling CatBoost</i>	33
3.3.8	Evaluasi.....	40
3.3.9	<i>Shapley Additive Explanations (SHAP)</i>	40
3.3.10	Analisis Hasil.....	45
BAB 4	46
HASIL DAN PEMBAHASAN	46
4.1	Hasil	46
4.2	Pembahasan.....	47
4.2.1	Pengumpulan Data	47
4.2.2	EDA	49
4.2.3	Preprocessing	57
4.2.4	Feature Engineering	62
4.2.5	Modelling <i>CatBoost</i>	66
4.2.6	Evaluasi.....	67
4.2.7	<i>Shapley Additive Explanations (SHAP)</i>	70
BAB 5	72
KESIMPULAN DAN SARAN	72
5.1	Kesimpulan	72
5.2	Saran	73
DAFTAR PUSTAKA	74
LAMPIRAN	77

DAFTAR TABEL

Tabel 2.1 Studi Literatur	10
Tabel 3.1 Statistik deskriptif	30
Tabel 3.2 labelling data	31
Tabel 3.3 Contoh data	34
Tabel 3.4 Contoh data kategorikal	39
Tabel 3.5 Hasil perubahan fitur kategorikal.....	39
Tabel 3.6 Contoh hasil prediksi model	41
Tabel 4.1 Fitur dataset.....	48
Tabel 4.2 Statistik deskriptif	50
Tabel 4.3 Rincian fitur dan jumlah <i>missing value</i>	58
Tabel 4.4 Fitur kategorikal sebelum dan sesudah penggabungan minoritas.....	58
Tabel 4.5 Fitur boolean sebelum dan sesudah perubahan tipe data	63
Tabel 4.6 Fitur kategorikal sebelum dan sesudah <i>one-hot encoding</i>	63
Tabel 4.7 Fitur sebelum dan sesudah <i>feature selection</i>	65
Tabel 4.8 Rincian <i>hyperparameter</i>	66
Tabel 4.9 <i>classification report</i>	68

DAFTAR GAMBAR

Gambar 2.1 Tampilan steam	17
Gambar 2.2 Machine Learning	18
Gambar 2.3 Cara Kerja Gradien Boosting	20
Gambar 2.4 Perbedaan pohon simetris dan tidak simetris	21
Gambar 2.5 Matriks Konfusi	22
Gambar 2.6 Kurva <i>AUC</i>	24
Gambar 3.1 Diagram Alir Penelitian	27
Gambar 3.2 Eksklusi gim indi.....	28
Gambar 3.3 Struktur data	29
Gambar 3.4 <i>Handle Missing Value</i>	30
Gambar 3.5 Penggabungan minoritas fitur kategorikal	31
Gambar 3.6 <i>Random Oversampling</i>	32
Gambar 3.7 Contoh penggunaan <i>one-hot encoding</i>	33
Gambar 3.8 <i>Feature Selection</i>	33
Gambar 3.9 Cara kerja Algoritma <i>CatBoost</i>	36
Gambar 3.10 Contoh penggunaan <i>CatBoost</i> pada data baru.....	38
Gambar 3.11 Interpretasi <i>SHAP</i> terhadap model	44
Gambar 4.1 games.json	48
Gambar 4.2 Struktur data	50
Gambar 4.3 Boxplot Fitur Numerik	52
Gambar 4.4 Distribusi kolom kategori 10 terbanyak	52
Gambar 4.5 Distribusi kolom genre 10 terbanyak	53
Gambar 4.6 Distribusi kolom bahasa 10 terbanyak	54
Gambar 4.7 Distribusi kolom bahasa audio 10 terbanyak	54
Gambar 4.8 Distribusi sistem operasi <i>windows</i>	55
Gambar 4.9 Distribusi sistem operasi <i>mac</i>	55
Gambar 4.10 Distribusi sistem operasi <i>linux</i>	56
Gambar 4.11 Barplot fitur <i>estimated_owners</i>	57
Gambar 4.12 Persentase dan jumlah label	61

Gambar 4.13	persentase dan jumlah data sesudah <i>random oversampling</i>	62
Gambar 4.14	Grafik <i>learn</i> dan <i>test error</i>	67
Gambar 4.15	Metrik konfusi data <i>testing</i>	68
Gambar 4.16	Kurva <i>AUC</i> pada <i>ROC</i>	69
Gambar 4.17	Interpretasi <i>SHAP</i>	70

DAFTAR LAMPIRAN

Lampiran 1. Data.....	77
Lampiran 2. Sintaks Scrape	77
Lampiran 3. Library	93
Lampiran 4. Exploratory Data Analysis.....	94
Lampiran 5. Preprocessing Data	98
Lampiran 6. Feature Engineering.....	100
Lampiran 7. Modelling CatBoost.....	104
Lampiran 8. Evaluasi	106
Lampiran 9. Shapley Additive Explanations	107